

Kinesin Superfamily Proteins (KIFs) in the Mouse Transcriptome

Harukata Miki,¹ Mitsutoshi Setou,¹ RIKEN GER Group² and GSL Members,^{3,4} and Nobutaka Hirokawa^{1,5}

¹Department of Cell Biology and Anatomy, Graduate School of Medicine, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; ²Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan; ³Genome Science Laboratory, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan

In the post genomic era where virtually all the genes and the proteins are known, an important task is to provide a comprehensive analysis of the expression of important classes of genes, such as those that are required for intracellular transport. We report the comprehensive analysis of the Kinesin Superfamily, which is the first and only large protein family whose constituents have been completely identified and confirmed *in silico* and at the cDNA, mRNA level. In FANTOM2, we have found 90 clones from 33 Kinesin Superfamily Protein (KIF) gene loci. The clones were analyzed in reference to sequence state, library of origin, detection methods, and alternative splicing. More than half of the representative transcriptional units (TU) were full length. The FANTOM2 library also contains novel splice variants previously unreported. We have compared and evaluated various protein classification tools and protein search methods using this data set. This report provides a foundation for future research of the intracellular transport along microtubules and proves the significance of intracellular transport protein transcripts as part of the transcriptome.

The mouse has been proven to be an excellent genetic model for the understanding of human biology. The availability of the genomic sequence of both organisms also allows for a comprehensive analysis of the catalog of classes of genes (Hattori et al. 2000; Kawai et al. 2001; Lander et al. 2001; Olivier et al. 2001; Venter et al. 2001; Waterston et al. 2002). In addition, a comprehensive analysis of the messenger-RNAs (mRNAs) produced in organisms (the transcriptome) was recently accomplished for the mouse (Okazaki et al. 2002), providing a global view of gene expression in this organism. This milestone is the beginning of a new era that will allow for the comprehensive analysis of systematic biology. The obvious next steps are to determine the function of genes. Cells transport and sort various proteins following synthesis as distinct kinds of membranous organelles and protein complexes to the correct destinations at appropriate velocities. This is true for all kinds of cells, both nonpolarized cells such as fibroblasts and polarized cells such as neurons and epithelial cells. Thus, intracellular transport is fundamental for cell morphogenesis, function, and survival (Hirokawa 1996, 1998).

The trafficking of proteins is tightly regulated and various different types of proteins are known to be involved. Members of the Kinesin Superfamily Proteins (KIFs) have been shown to transport organelles, protein complexes, and mRNAs to specific destinations in a microtubule- and ATP-dependent manner (Hirokawa 1996, 1998; Brendza et al. 2000). KIFs also participate in chromosomal and spindle

movements during mitosis and meiosis (Vale and Fletterick 1997; Hirokawa et al. 1998; Sharp et al. 2000). KIFs contain a motor domain region, highly conserved among all eukaryotic phyla studied thus far, that includes a p-loop motif, switch 1 and 2 motifs, and microtubule-binding regions (Vale and Fletterick 1997; Hirokawa 1998; Kim and Endow 2000; Kikkawa et al. 2001). Microtubules serve as rails for these transportation proteins and have a polarity in a manner in which there is a fast-growing plus end and a relatively stationary minus end. The organization is tightly regulated in cells. In nerve axons, microtubules are arranged longitudinally with the plus end oriented away from the cell body. In proximal dendrites the polarity of microtubules is mixed, whereas at the distal end, the polarity is the same as that in the axon. In epithelial cells microtubules are organized with the plus end oriented toward the basement membrane. In most other cells such as fibroblasts, microtubules radiate from the cell center with the plus end oriented toward the periphery.

KIFs can be divided into three classes depending on the location of the motor domain in the molecule. N kinesins and M kinesins, containing motor domains close to their N terminal or center, have been reported to possess microtubule plus end-directed motility. There are three KIFs containing a motor domain proximal to the C terminus and possessing minus end-directed motility. Microtubule plus end-directed transport is mainly driven by KIFs, whereas cytoplasmic dynein is responsible for the bulk of microtubule minus end-directed transport. The Kinesin Superfamily is the first and only large protein family whose constituents have been completely identified and confirmed *in silico* and at the cDNA or mRNA level (Miki et al. 2001). Analysis of KIFs is an efficient way to assess the functional protein content of a library and our report is an example of the possibilities provided by the

⁴Takahiro Arakawa,² Piero Carninci,^{2,3} Jon Kawai,^{2,3} and Yoshihide Hayashizaki^{2,3}

⁵Corresponding author.

E-MAIL hirokawa@m.u-tokyo.ac.jp; FAX +81-3-5689-4856.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.984503>.

FANTOM2 clone set for the analysis of a complete protein family.

To set the foundation for functional genomics of intracellular transport network in the transcriptome, we have analyzed the Kinesin Superfamily, an essential component of the microtubule (MT)-dependent transport system in the largest cDNA library to date, the FANTOM2 library.

RESULTS

KIF Clones in FANTOM2

Of the 45 KIF loci identified in the genome, representative transcripts of 33 loci were found in the FANTOM2 library (Table 1). The 33 representative sequences arise from a total of 90 clones deriving from 49 libraries.

Seventeen representative transcripts were full length (51.5%); two sequences had problems other than truncation (6.1%), specifically, one had a 1.5-kb deletion in the middle of the coding region and one locus was represented by an unspliced genomic fragment (Fig. 1). Seven representative transcripts were 3' truncated (21.2%) and six were 5' truncated (18.2%). One representative clone was 5' and 3' truncated. Twenty out of the 90 KIF clones did not contain the signature motor domain motif.

KIF Clones in Phase I

KIF transcripts not found in the FANTOM2 data set were searched for in the Phase I set. The 12 KIFs with no transcripts in the FANTOM2 data set are KIF1A, KIF4B, KIF5A, KIF8, KIF10, KIF16B, KIF19A, KIF19B, KIF26A, KIF26B, KIFC2, and KIFC3. As a result of BLASTN searches using nucleic acid sequences, ESTs of 9 KIFs (KIF1A, KIF5A, KIF10, KIF16B, KIF19A, KIF26A, KIF26B, KIFC2 and KIFC3) were identified. Excluding KIF16B, the eight other KIFs had ESTs deriving from tissue abundant in neurons. ESTs of KIF4B, KIF8, and KIF19B were not detected in any EST database examined.

Detection in Neural Tissue

In FANTOM2, 25 loci had clones coming from neural or mixtures of neural and other tissue (Fig. 2). These clones derived from libraries

Table 1. KIF Clones Found in Fantom 2

Locus	Clone ID	Library	Sequence state	
KIF1B (5)	3110082B06	E13 head	3'UTR	
	9630038C03	N16 cerebellum	(-) 3'	
	A530055P03	Aorta and vein	3'UTR	
	A530096N05	Aorta and vein	(-) 5', no motor	
	D830006L05	N16 heart	(-) 5', no motor	
KIF1C (2)	E130317N11	N0 eyeball	(-) 3', no motor	
	4831440N04	N0 head	(-) 3'	
	B430105J22	Adipose	(-) 5', 3'	
KIF2A (3)	4933407G21	Testis	(-) 3'	
	9930101N02	Vagina	(-) 5', 3'	
	C530030B14	E12 spinal cord	Genomic, no motor	
KIF2B (1)	4933413L02	Testis	Full length	
KIF2C (1)	E430002N04	N2 thymic cells	Full length	
KIF3A (1)	6030448M04	E13 testis	(-) 3', no motor	
KIF3B (2)	C130035P16	E16 head	(-) 3', variant	
	D030068F10	E9	Full length	
KIF3C (4)	4833419C04	N0 head	(-) 5'	
	9530058P10	U. bladder	3' UTR	
	C630047J11	Hippocampus	3' UTR	
	G630016D04	Mixed	3' UTR	
	D330050K22	E13 heart	Full length	
KIF4A (1)	4632419K17	N0 skin	(-) 5'	
KIF5B (5)	6030497E13	E13 testis	R&C, (-) 3', no motor	
	9430093P06	E12 upper body	(-) 3' no motor	
	A130087D15	N16 thymus	3' UTR	
	D530024B08	E12 stomach	(-) 5', 3'	
	A430099A21	N0 thymus	R&C, (-) 3'	
KIF5C (2)	A730049O10	N7 cerebellum	(-) 5', 3' no motor	
	A730092G14	N7 cerebellum	(-) 3'	
KIF6 (2)	D130004B10	E12 spinal ganglion	Full length	
	4930404H06	Testis	R&C	
KIF7 (3)	9330171B17	Diencephalon	R&C	
	A230018M15	Hypothalamus	R&C, full length	
	1700062H04	Testis	(-) 5', 3'	
KIF9 (5)	4930548D19	Testis	(-) 3'	
	4933417D18	Testis	(-) 5'	
	4921509F14	Testis	Full length	
	E030019L05	N0 lung	Full length	
	C920004O08	N2 thymic cells	Full length	
KIF11 (2)	D030019B03	E9	(-) 5', (+) introns	
	9130007B22	Cecum	Full length	
KIF12 (1)	4932439H05	Testis	(-) 5', 3'	
	4930505I07	Testis	(-) 5'	
KIF13A (4)	9930023M01	Vagina	(-) 5'	
	A330065I17	Spinal cord	(-) 5', no motor	
	A930029L02	Retina	Full length	
	E130203M01	N0 eyeball	Variant	
	D030022H04	E9	(-) 3'	
KIF13B (1)	D330038N01	E13 heart	(-) 3'	
	7030401O13	E13 gonad	(-) 5', 3'	
KIF14 (1)	5930435E01	E8 forelimb	Variant	
	B930001E07	N10 cerebellum	(-) 3'	
KIF15 (2)	C630023I21	Hippocampus	(-) 5', problem	
	9830166P06	Bone	(-) 3'	
	A630013A09	N3 thymus	(-) 3'	
	B130001M12	Parthenogenetic embryo	(-) 3'	
	C330012H11	ES cells	Full length	
KIF16A (1)	C430017A15	E15 whole	(-) 3'	
	4921519A02	Testis	Full length	
KIF17 (3)	D030002O10	E9	Full length	
	D030068C15	E9	Full length	
	D130015N09	E12 spinal ganglion	Full length	
	D230032N19	E12 eyeball	Full length	
	D330036L06	E13 heart	Full length	
	E130006J15	N0 eyeball	Full length	
	C330014J10	ES cells	(-) 3'	
	B130024C23	Parthenogenetic embryo	(-) 3'	
	KIF18B (1)			
KIF20A (6)				
KIF20B (2)				

(continued)

Table 1. *Continued*

Locus	Clone ID	Library	Sequence state
KIF21A (8)	2010012N14	Small intestine	(-) 3'
	4833440H09	N0 head	(-) 3'
	B930036B02	N10 cerebellum	(-) 5', no motor
	B930052C11	N10 cerebellum	(-) 3'
	C130019F07	E16 head	(-) 5', 3'
	C330021D17	ES cells	(-) 5' no motor
	D730038E15	L10 mammary gland	(-) 5' no motor
	E130012K01	N0 eyeball	(-) 5' no motor
KIF21B (2)	C230056K15	N0 cerebellum	(-) 5' no motor
	F730023C13	B6-derived CD11+ve dendritic cells	(-) 5', 3'
KIF22 (4)	2400004I16	ES cells	Full length
	B130006E01	Parthenogenetic embryo	(-) 5', (+) intron
	D230020I05	E12 eyeball	Full length
	E430002J24	N2 thymic cells	(-) 5'
KIF23 (3)	310001D19	E13 head	(-) 3'
	4632406J10	N0 skin	(-) 5'
	5730589H11	E8	Full length
KIF24 (3)	4933425J19	Testis	(-) 3', (+) intron
	D030003D17	E9	(-) 3'
	D430019P19	E13 lung	(-)2230-3868
KIFC1 (2)	9230115E21_	Extra testis	R&C, problem
	C130020G17	E16 head	(-) 5'

(No. of clones)

made from hippocampus, hypothalamus, spinal cord, retina, and mixed libraries such as whole embryos. Adding into consideration the KIFs found in the Phase I set, transcripts of 33 KIFs were found in nervous tissue or body parts containing neural tissue such as sensory organs. One clone encoding the 5' and 3' end of KIF1A derived from a diencephalon library. One 3' end sequence from a clone found in a 16-dpc (days post conceptus) embryo head library is identical to part of KIF5A. ESTs matching KIF10 originated in spinal cord and eyeball libraries, KIF19A in inner ear, KIF26A in whole embryos, KIF26B in neonate head, KIFC2 in diencephalon and other neuronal tissues, and KIFC3 in embryonic head. Four KIF16B ESTs were found but none in neuronal tissue.

Alternative Splicing

Two previously reported isoforms deriving from the KIF1B loci were identified in FANTOM2 (Nangaku et al. 1994; Zhao et al. 2001). Variants are indicated in the "sequence state" column of Table 1. Four KIFs, namely, KIF3B, KIF9, KIF17, and KIF24, had alternative splice variants not reported previously (Table 1; Fig. 3).

Comparing the two KIF3B transcripts, C130035P16 and D030068F10, the former is an unreported isoform and the latter is identical to the sequence deposited in GenBank (NM_008444). The new isoform has only two exons. The first exon is shared until base 1095. There the novel form splices and connects to the second exon, which is unique to the variant and is located in the genome between the sixth and seventh exons of the conventional form. The intron between the first and second exon starts with the nucleic sequence GA and ends with AG. Twenty-three ESTs in the public database specifically support the previously reported form whereas one EST is specific for the novel form. The open reading frame (ORF) of the original form translates into 747 amino acids, the new form into 329 residues, excluding a one-base insertion

that is not supported by the original clone nor the genomic sequence.

Clone E030019L05 is identical to the KIF9 sequence in GenBank (NM_010628). Clone 4921509F14 is identical until the 774th amino acid residue, after which the conventional KIF9 sequence has 16 residues whereas the novel one has a different 36 residues. The two isoforms share the first 17 exons. The conventional form has an 18th and 19th exon, which are located downstream of the last exon of the variant in the genome. Eleven ESTs in the NCBI mouse EST database support the original isoform; 29 support the novel isoform.

KIF17 has a previously unpublished variant, 5930435E01. This splice form lacks the 8th, 9th, and 15th exons of the published form (accession no. AB008867). As a result, the first 940 amino acids are shared excluding residues 411–649. Because of a frame-shift resulting from the deletion of the 15th exon,

the last 8 amino acid residues are specific to the novel isoform. The presence of the 8th and 9th exons are supported by 2 ESTs and the 15th exon is supported by an additional 2 ESTs. There is one EST lacking the 15th exon deposited.

In the FANTOM2 library, there are three KIF24 clones, all of which contain different sequences resulting from alternative splicing. Clone 430019P19, the longest of the three, is encoded by 10 exons. Clone D030003D17 ends in the 8th exon of clone 430019P19 without any in-frame stop codon but contains four exons between the 3rd and 4th exons of the former clone. Clone 4933425J19 shares the first seven exons with clone D030003D17. However, the 7th exon is extended beyond the splice site for D030003D17 and yields an in-frame stop codon. Clone 4933425J19 contains three separate bases dispersed throughout the transcript not found in the other two clones nor in the genome and not considered in this study. The 3' end of the longest clone, 4933425J19, matches eight ESTs in the public database. One EST supports the four exons included in clone 4933425J19; in contrast, no EST was found that agreed with D030003D17 in leaving out the four exons. Clone D030003D17 encodes an 862-amino-acid protein, 430019P19, 747 residues and 4933425J19, 371 residues, ignoring the 3 base insertion.

Identification of KIF Clones

Of the FANTOM2 clones, 57 were defined as KIFs by Pfam, 53 by InterPro, 68 by Gene Ontology, 102 by auto-annotation, and 81 by FANTOM2 annotation (Fig. 4). Of these sequences, InterPro defined 1 false clone, and Gene Ontology, auto-annotation, and annotation defined 7, 18, and 8 false clones, respectively. Twenty-eight KIFs were found by all five methods, 22 were found by 4 methods or 3 methods, and 15 were found by 2 methods. One clone was singly identified by annotation and 2 were singly identified by auto-annotation. In-

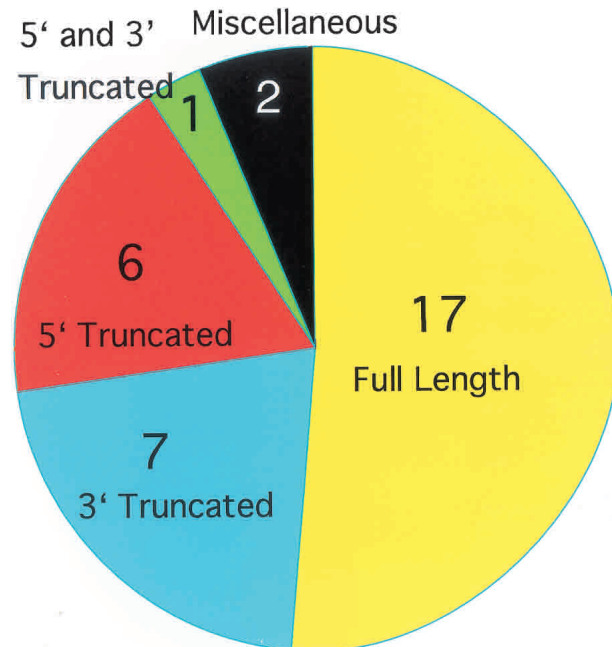


Figure 1 Coverage of KIFs in FANTOM2. Of the 45 KIF loci found in the genome, 33 representative transcripts were found in FANTOM2, of which, 17 (51.5%) had full-length clones. Seven (21.2%) loci had 3' truncated clones, 6 (18.2%) had 5' truncated clones, 1 (3.0%) had 5' and 3' truncation, and 2 (6.1%) had clones with other problems.

terPro mis-selected 1 false positive, Gene Ontology 7, auto-annotation 18, and annotation 8. Pfam did not recognize any false KIFs.

BLASTN and TBLASTN searches using the nucleotide and amino acid sequences of KIFs did not reveal any further clones in the FANTOM2 set.

Phylogeny of KIFs in FANTOM2

KIFs affiliated with 13 out of 14 classes were represented by at least one gene in FANTOM2 (Fig. 5). Seven classes of KIFs out of 14 had all members represented in FANTOM2, including all orphan KIFs. These classes are class N-4 kinesins, N-6, N-9, N-10, M, and C-1. Orphan KIFs refer to KIF6, KIF7, and KIF9. These KIFs do not have any orthologs in *Drosophila melanogaster*, *Caenorhabditis elegans*, or *Saccharomyces cerevisiae*. Ten subfamilies out of a total of 18 had all members included in FANTOM2: the KIF2, KIF3, KIF12, KIF13, KIF15, Osm 3/KIF17, KIF18, Rab 6-KIF/KIF20, MKLP 1/CHO 1, and NCD/Kar 3 subfamilies. Concerning full-length coverage, all orphan members and all constituents of four subfamilies were represented by full-length clones, namely, the KIF12, Osm 3/KIF17, KIF18, and MKLP 1/CHO 1 subfamilies.

When including KIFs found in Phase I, all classes and subfamilies are represented.

DISCUSSION

Two sets of molecular motors, KIFs and dyneins, use the microtubule cytoskeleton as rails. Of the 45 KIF loci in mouse, representative transcripts from 33 loci were found in FANTOM2 along with 5 novel isoforms. Adding the 2 isoforms of KIF1B, the resulting TU coverage for KIFs in FANTOM2 is 86.7%. When considering the Phase I clones, the coverage rises to 94.1%, both values in good agreement with

the overall FANTOM2 TU coverage of 90.1%. Twelve KIFs were not found in FANTOM2. The lack of KIFs normally abundant in other cDNA libraries may reflect the thorough subtraction of abundant transcripts conducted during the development of FANTOM2. Despite subtraction, 25 KIFs out of 33 found in FANTOM2, equivalent to 75.8%, derived from neural tissue or mixtures of neural and other tissue. Including sequences found in the Phase I set, the percentage is 78.6%. Previously, we have reported that a similar percentage, 84.4%, of the KIFs (38 out of 45) have been detected in neural tissue (Miki et al. 2001). Six KIFs previously found in neural tissue were not found in brain or other neural tissue. One KIF, KIF24, which was not detected in adult neural tissue previously, was found in a 9-dpc whole embryo library. The tissues where KIF24 is expressed in the embryo are yet to be determined, but this is the first time it has been detected in embryo. It should be kept in mind that the derived library in FANTOM2 may not be the only tissue in which the clone has been detected. The clones have been through a screen for unique sequences and, therefore, if several clones from different libraries have the same sequence, only one sequence would be selected to represent all exactly matching clones.

The Phase I data set is comprised of 547,149 5' end sequences and 1,442,236 3' end sequences collected to select clones with unique 5' and 3' end sequences for the FANTOM2 clone set (Okazaki et al. 2002). These end sequences of clones are deposited as ESTs. Therefore, the complete length and full sequence of clones containing these sequences are unknown. Additionally, these clones are not available for distribution. The purpose of using the Phase I set for this study is twofold: first, to identify KIFs not found in the FANTOM2 set and, second, to determine which library those clones originated in. In some cases, EST information included in the Phase I set was used to validate newly identified alternative splice variants found in FANTOM2. Twelve KIFs were not found in the FANTOM2 set, including abundantly expressed KIFs such as KIF1A (Okada et al. 1995) and KIF5A (Aizawa et al. 1992). Nine KIFs out of the missing 12 were found in Phase I (Table 2). The three KIFs that could not be identified in the Phase I database were KIF4B, KIF8, and KIF19B. It should be noted that KIF25 is not found in the mouse genome nor in any mouse cDNA library including FANTOM2 and Phase I. However, it is present in human genome databases and there are abundant human ESTs (Okamoto et al. 1998; this study). This KIF was also searched for in the FANTOM2 and Phase I databases but could not be found. Many of the genes in the proximity of the KIF25 locus in the human genome are absent in the corresponding region in the mouse genome. It is possible that during evolution, after the separation of mouse and humans, humans acquired or mice lost the genomic region close to the KIF25 locus. The precise function of KIF25 in cells is currently not known (Okamoto et al. 1998). Gene knock-out studies using mice have identified KIFs that can be deleted from the genome without creating a detectable phenotype (Yang et al. 2001; Nakajima et al. 2002). Cells may possess a compensatory function for these KIFs. Alternatively, KIF25 may contribute to the difference in humans and mice. The increased complexity of human biology may demand a compatible intracellular transport system. There is a possibility that it exists in the mouse genome in a region not yet fully sequenced and simultaneously rarely expressed, making it difficult to sequence ESTs. KIF4B was not detected as an EST in this database nor has it been found in any other library, though it has a locus in the mouse genome (Ha et al. 2000;

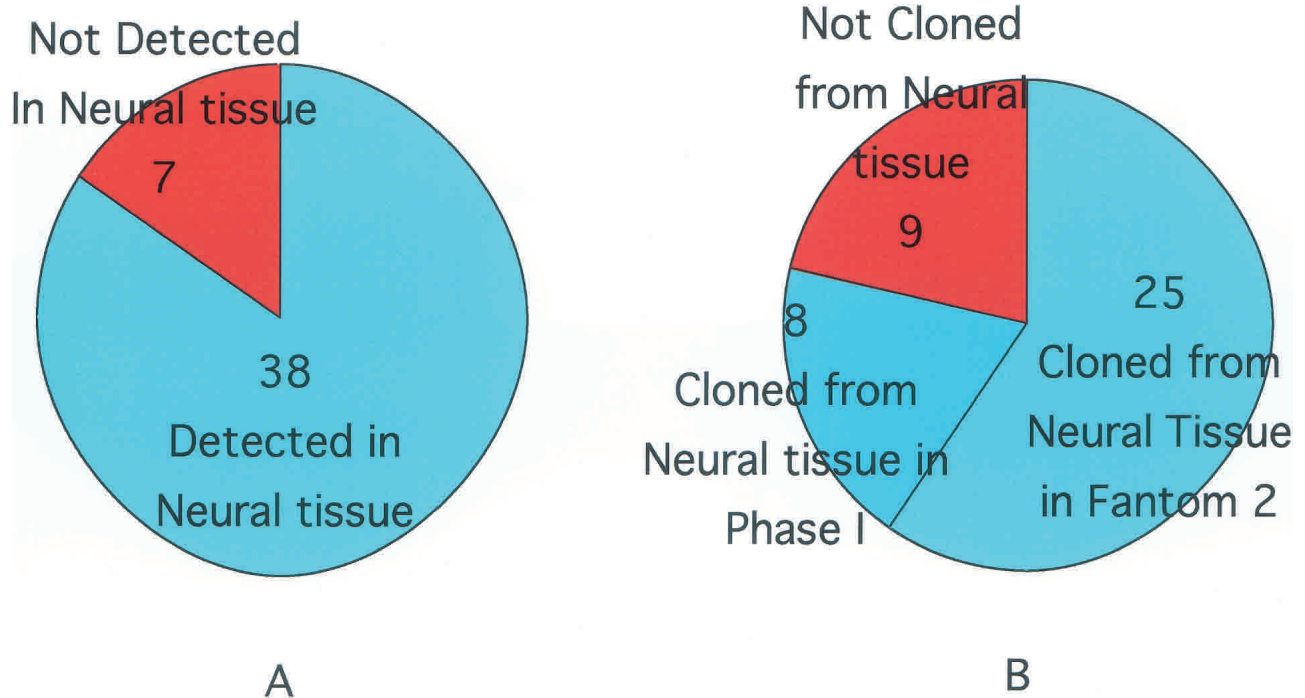


Figure 2 Consistency of KIF detection in neural tissue. (A) Previously, 38 KIF transcripts (84.4%) were detected in brain or other neural tissue. (B) In FANTOM2, 25 (75.8%) derived from neural tissue or mixtures of neural and other tissue. Adding the Phase I clones, 33 (78.6%) out of 42 were identified in neural tissues.

Miki et al. 2001). Sequences in the locus and the predicted transcript are over 83% homologous to KIF4A. The high homology suggests that one of the two genes has arisen through gene duplication. KIF4B may be expressed, albeit at levels so extremely low as to be undetectable. It is also plausible that it is not expressed as it has never been detected as a transcript even after extensive searches in which all other KIF transcripts have been identified (Miki et al. 2001). The locus of KIF8 is yet to be identified in human and in mice though it has been detected by PCR in a mouse cDNA library (Nakagawa et al 1997). It may be in the genome in a region not yet fully sequenced and simultaneously rarely expressed. The KIF19B locus has been located and cDNA identified previously (Miki et al. 2001). Three KIFs have not been detected in the FANTOM2 library, of which only the transcript of KIF4B has not been found and reported in any other library. The completeness of the number of KIFs in the FANTOM2 and Phase I databases is impressive.

There are two previously reported isoforms of KIF1B (Nangaku et al. 1994; Zhao et al. 2001). Both were identified in the FANTOM2 clone set. The KIF3B (Yamazaki et al. 1995), KIF17 (Setou et al. 2000), and KIF24 (Miki et al. 2001) clones contained novel 3' sequence revealing previously unknown splice variants. The novel KIF3B variant would encode a protein that contains an intact motor domain, the functional domain of KIFs comprised from ATP-binding and microtubule-binding motifs. The protein terminates one amino acid after the 7th alpha helix, the end of the motor domain. This indicates the splice variant is motile and functional. The amino acid sequence of the KIF9 splice variant has a longer and different COOH terminal, implying it would bind to alternative proteins. The COOH terminal of the conventional isoform binds to a GTPase (Piddini et al. 2001). The fact that

there are more than twice the number of ESTs for the novel form indicates it is expressed at a higher level than the previously reported form. The distribution of clones is representative of the results of Northern blotting, suggesting that expression levels are reflected to which libraries the sequences are found in. The new KIF17 isoform lacks the 2nd microtubule binding site along with Switch 1 and 2 ATP-binding domains, suggesting it is not processive. Previous reports have used dominant negative forms of KIFs by expressing tail sequences lacking motile function (Nakagawa et al. 2000; Setou et al. 2002; Guillaud et al. 2003). The novel isoform presumably would have a similar function, thereby regulating intracellular transport. Of the three KIF24 isoforms, only D030003D17 has a functional motor domain as predicted by presence of all ATP-binding and microtubule-binding motifs. Though the sole method of proving that a motor is motile is to do a motility assay, intact ATP-binding and microtubule-binding motifs are required for motility (Kikkawa et al. 2001). These new isoforms demonstrate the diversity and depth of molecular representation in the transcriptome and add a novel diversity to the Kinesin Superfamily. This finding is significant because compared with the innumerable molecules transported within the cell, the relatively limited number of KIFs imply a complex transport mechanism involving multiple splice variants and adaptor proteins.

The KIF motor domain is comprised of highly conserved ATP-binding and microtubule-binding motifs, which are required for motility. The p-loop binds ATP, whereas switch 1 and switch 2 form a salt bridge that is broken upon release of γ -phosphate from ATP. The collapse of the salt bridge alters the conformation of the protein, resulting in movement (Kikkawa et al. 2001). The amino acid sequences of the p-loop, switch 1, and switch 2 can be characterized as "GXXXXGK(S/

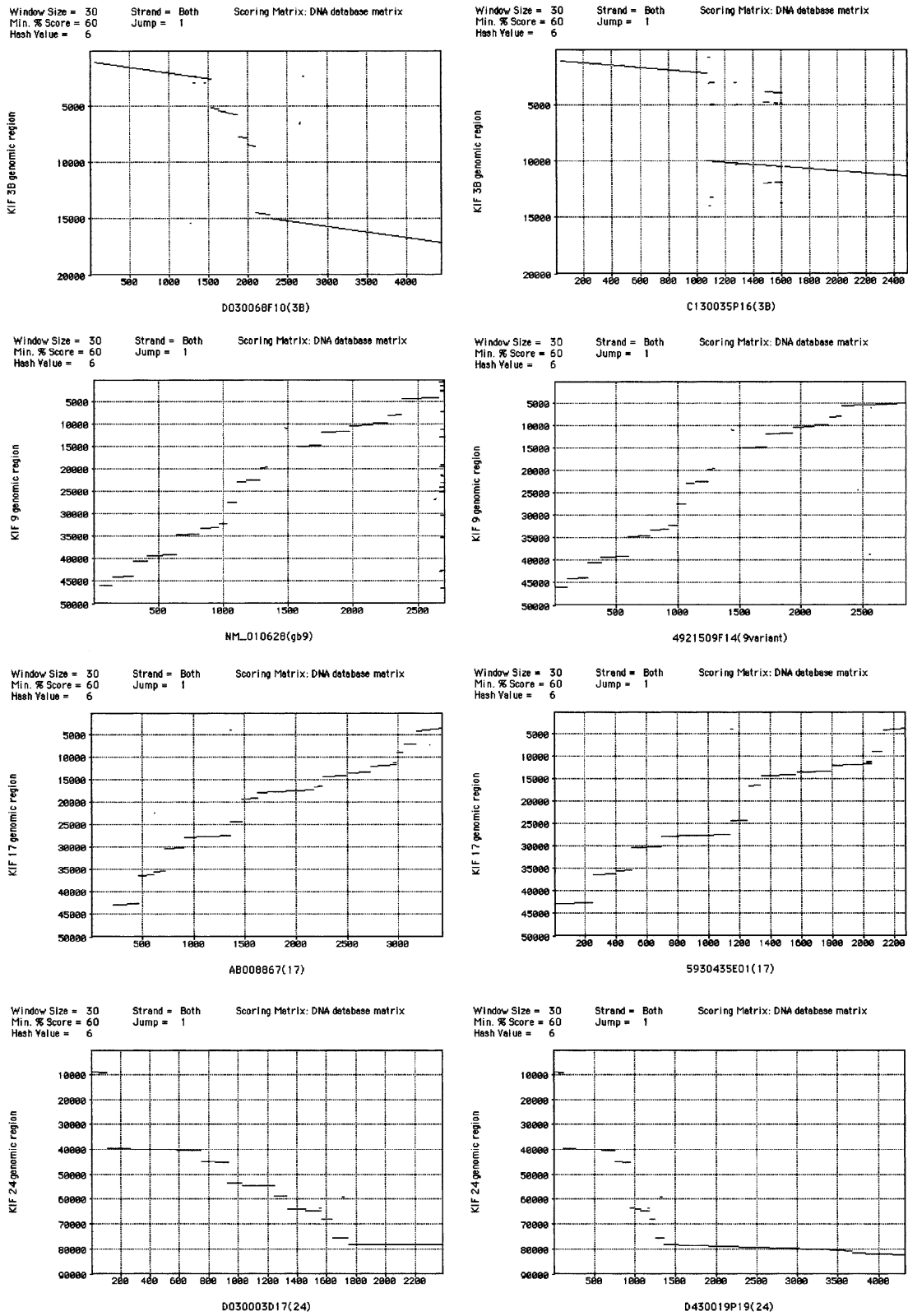


Figure 3 Alignment of transcripts and genomic sequences. Transcribed sequences of KIF3B, KIF9, KIF17, and KIF24 were aligned with respective genomic sequences to reveal intron–exon structures.

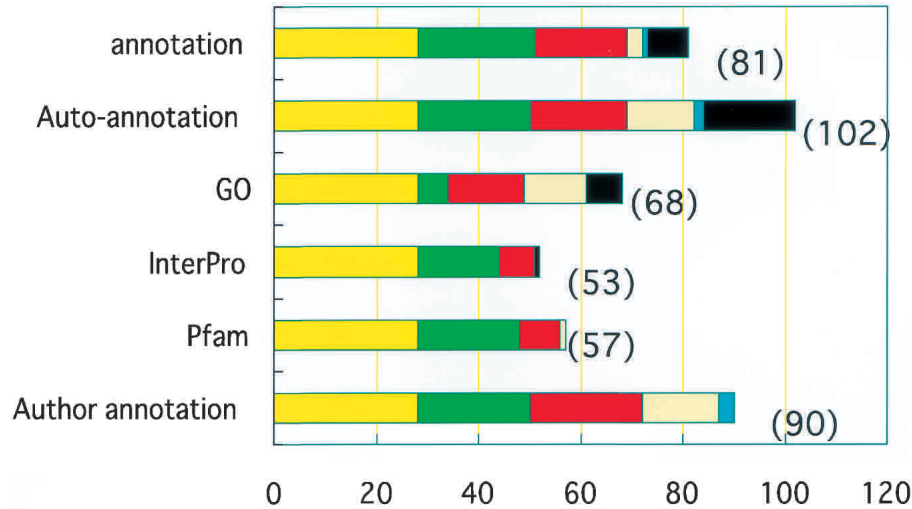


Figure 4 Protein search tool comparison. Five methods of detecting KIFs were compared. Twenty-eight clones were detected by all five methods. Pfam and InterPro had low false positive and high false negative rates. Auto-annotation detected the most KIFs but also the most false positives. The false positives were greatly reduced from 18 to 8 by human annotation. Clones identified by respective number of search tools are indicated by the following colors: (yellow) all 5 search tools, (green) 4 search tools; (red) 3 search tools; (white) 2 search tools; (blue) 1 search tool; (black) false positive.

T)", "SSRSH", and "DLAGSE", respectively, where X represents any amino acid. The microtubule-binding site 3, though less conserved compared with the ATP-binding motifs, can be represented by the amino acid sequence "HVPYRD" downstream of switch 2. It is perceived that these motifs must all be present and in this order. Thus, these sequences were used for the manual identification of KIFs in this study and previously by us and other researchers (Aizawa et al. 1992; Nakagawa et al. 1997; Yang et al. 1997; Miki et al. 2001; Reddy and Day 2001).

Of the 57 KIFs identified by Pfam, 28 were identified by all other methods. These numbers implicate the accuracy of Pfam in identifying KIFs. However, Pfam did not succeed in detecting 33 clones, including 12 clones that contain an intact switch 2 consensus sequence, which is used for identifying KIFs. The reason these 12 clones were not selected cannot be inferred from the amino acid sequence. There is a possibility that the algorithm used can be improved to increase sensitivity. InterPro hits contained one kinesin light chain clone that comprises a separate category. InterPro uses several criteria including p-loop, switch 1, and switch 2 sequences. These two protein motif search engines had a very low false positive rate but had a high false negative rate even for clones containing an intact signature motif. Pfam and InterPro search and identify protein motifs. These motifs are then used to classify proteins by Gene Ontology. Gene Ontology classification categorized seven clones falsely as KIFs but was successful in detecting more KIF sequences than the aforementioned two motif search engines. Auto-annotation picked up the most false positives, detecting 18 false KIFs. These hits include kinesin light chains and GenBank entries containing words such as "similar to Rab6 kinesin". These 18 false auto-annotations were decreased to 8 by human annotation. The decrease in false positives and high detection rate infer the necessity of human curation. These two methods correctly choose 84 and 72 clones out of ninety, respectively. Auto-annotation missed two full-length clones with identical se-

quences deposited in GenBank. By human annotation, no clone containing the signature motif was neglected. All other clones that were not selected by the two methods, the false negatives, only contained UTR sequences or needed to be reversed and complemented or lacked exons existing in the GenBank sequence. These clones are difficult to identify unless thoroughly familiar with various KIF sequences. Therefore, to identify pre-existing KIFs deposited in databases such as GenBank, human annotation may be the best method having a high detection rate, the advantage of not requiring motor domain consensus sequences in truncated clones, and the reduction of false positives by human curation. However, protein motif search engines may categorize better new full-length clones not previously deposited in any database where there would be no exactly matching reference. False positives can be

reduced by the exclusion of kinesin light chains that do not contain motor domains and GenBank deposit sequences that are titled "similar to kinesin," etc.

Twenty out of the 90 KIF clones did not contain the signature motif, equivalent to 22.2%. This percentage is similar to the 25% lower protein motifs found in the over all CDS in FANTOM2. More clones contained the full-length sequence than not. The percentage of 5'-truncated and 3'-truncated clones were approximately equal, indicating the possibility that there is no preference in truncation of either end. Only one locus was represented by a 5' and 3' truncation, adding evidence to the quality of this clone set. Only one locus was represented by a clone with other problems. Although some of the clones not containing the motor domain may have become truncated during reverse transcription or other technical steps, it is possible that these clones exist in vivo. As described for the KIF17 splice variant above, these transcripts would function as dominant negative regulators of cargo binding. Intracellular transport of cargoes could be controlled by competitive binding of cargo binding domains of intact and truncated KIFs. The 3' truncations also may be due to technicalities, though the possibility they exist in vivo cannot be denied. These transcripts would function in the cell as a result of transcriptional regulation and/or may bind alternative binding partners by exposing domains that are hidden in longer transcripts.

Seven out of 14 classes of KIFs and 10 out of 18 subfamilies had clones from all loci. This is a high number for a cDNA library and reflects the high coverage of TUs in FANTOM2. In addition, all members of five subfamilies were represented by full-length clones.

The KIFs reflect the representation of the transcriptome in FANTOM2, and this representation is in good agreement with predictions of all transcripts. It is highly possible that the predictions are accurate as indicated by the highly similar indicators of the KIFs. The high occurrence of the KIFs and the abundance of full-length clones implicate the necessity of us-

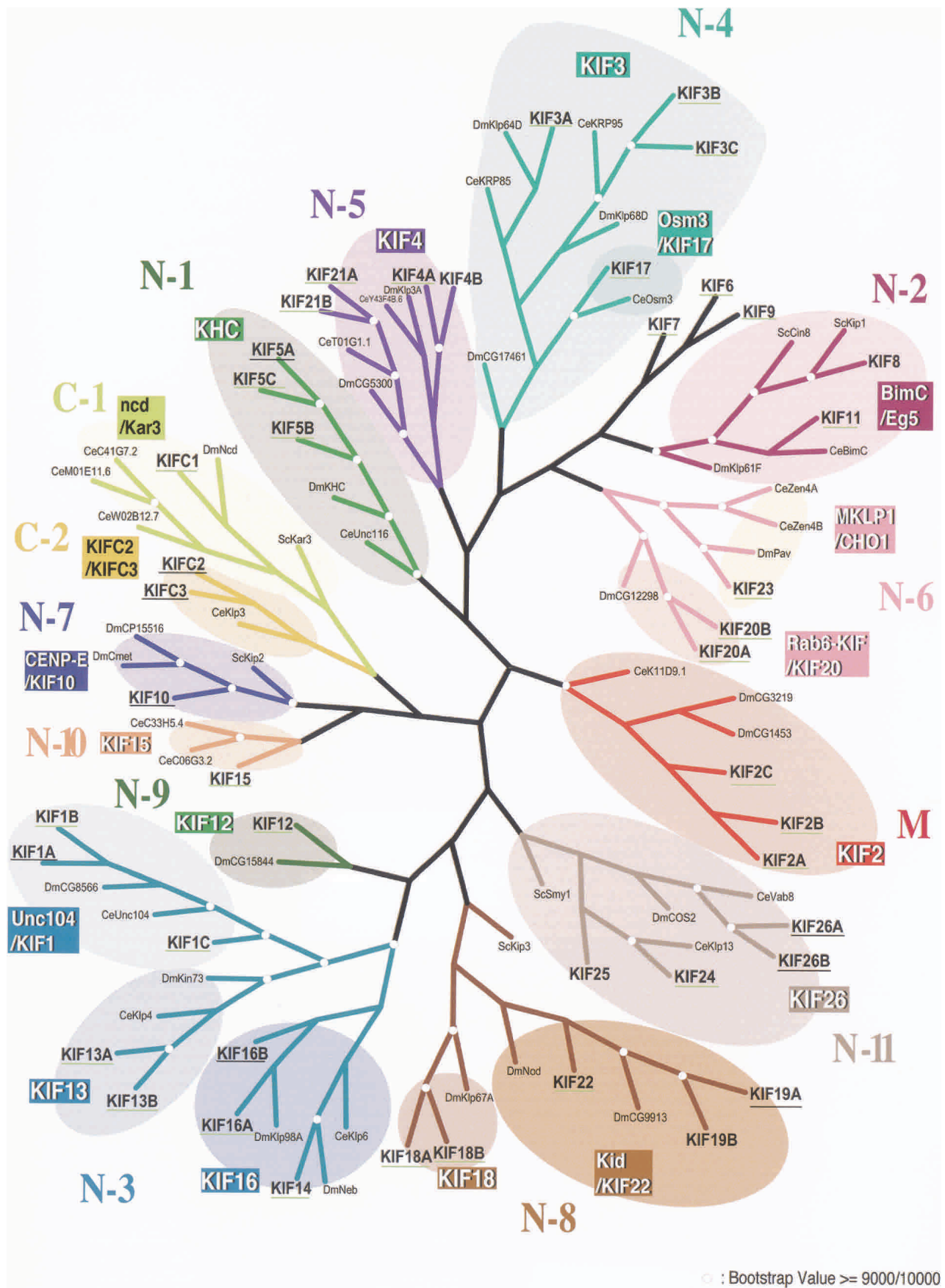


Figure 5 Phylogenetic tree of all KIFs found in mouse and human, flies, nematodes, and yeast. KIFs affiliated with 13 out of 14 subfamilies were represented by at least one gene in FANTOM2. KIFs found in FANTOM2 are underlined in yellow-green. Transcripts found in Phase I are underlined in black.

Table 2. ESTs of KIFs in the Phase I Data Set

Locus	Position = EST
KIF1A	5' = BB625880 diencephalon, clone 9330140A05 5' 3' = BB077437 diencephalon, clone 9330140A05 3'
KIF5A	1028 – 1235(/3639) = BB362667 16dpc embryo head, C130012K23 3'
KIF10	2kb = BB652685 12 dpc embryo spinal cord, C530022J18 5' 3' = BB420480 12 dpc embryo spinal cord, C530022J18 3' 3' = BB701713 in vitro fertilized eggs, 7420432D06 3' 3' = BB496665 0 day neonate kidney, D630004C19 3' (95%) 3' = BB473232 12 dpc embryo eyeball, D230050E14 3' (94%) 3' = BB439577 9 dpc embryo Mus musculus, D030018D11 3' (94%)
KIF16B	5' = BB605439 0dpc lung, E030011D06 5' (93%) 5' = BB593935 4dpc adipose, B430304K12 5' (93%) 3' = BB666006 oviduct, E230025N21 5' (97%) 3' = BB530665 0dpc lung, E030011D06 3' (92%)
KIF19A	5' = BB850223 inner ear, F930105P20 5' (98%) 5' = BB849042 inner ear, F930006C05 5' (96%)
KIF26A	3' = BB733390 12 dpc embryo whole body, E970038J21 3' 3' = BB328554 4 days neonate male adipose, B430319I09 3' (94%) 3' = BB072180 colon, 9030008O16 3' (96%)
KIF26B	2576 = 0 day neonate head, 4832420M10 5' 3' = 0 day neonate head, 4832420M10 3' (95%) 3' = B16 F10Y cells, G370079H03 3' 3' = medulla oblongata, 6330508I11 3'
KIFC2	5' = BB871391, 16 days neonate male diencephalon, G630029D09 5' 5' = BB869387, pooled tissues, intestinal mucosa, etc., G630012I08 5' 5' = BB870117, 16 days neonate male medulla oblongata, G630018H14 5' 5' = BB870711, 16 days neonate male medulla oblongata, G630024J17 5' 5' = BB868041, pooled tissues, intestinal mucosa, etc., G630001N07 5' 2090 – 2731 = 10 days neonate cortex, A830006K07 3' 3' = BB770313, B16 F10Y cells, G370107L04 3' 3' = BB799871, 16 days neonate male diencephalon, G630026D11 3' 3' = BB799203, 16 days neonate male diencephalon, G630022C05 3' 3' = BB799409, 16 days neonate male diencephalon, G630023D13 3' 3' = BB798223, 16 days neonate male diencephalon, G630012I06 3' 3' = BB799096, pooled tissues, intestinal mucosa, etc., G630021C17 3' 3' = BB802548, 16 days neonate male diencephalon, G630047I09 3' 3' = BB799651, 16 days neonate male medulla oblongata, G630024J17 3' 3' = BB807291, 16 days neonate male diencephalon, G630080J12 3' 3' = BB803692, 16 days neonate male diencephalon, G630057H09 3' 3' = BB808936, 16 days neonate male diencephalon, G630092J11 3' 3' = BB807614, 16 days neonate male diencephalon, G630082P18 3' 3' = BB804622, 16 days neonate male diencephalon, G630065D07 3' 3' = BB804925, 16 days neonate male diencephalon, G630066J23 3' 3' = BB798915, 16 days neonate male diencephalon, G630019L13 3' 3' = BB804241, 16 days neonate male diencephalon, G630062N06 3' 3' = BB798750, 16 days neonate male medulla oblongata, G630018H14 3' 3' = BB809333, 16 days neonate male diencephalon, G630095G10 3' 3' = BB802826, 16 days neonate male diencephalon, G630050G12 3' 3' = BB799720, 16 days neonate male diencephalon, G630025A16 3' 3' = BB797789, 16 days neonate male diencephalon, G630008D09 3' 3' = BB807174, 16 days neonate male diencephalon, G630079J18 3'
KIFC3	2585 – 2869 = 16 days neonate male medulla oblongata, G630071I23 5' = BB615386, adult male testis 4930502J08 5' 925 – = BB584419, adult male epididymis, 9230105K18 5' 1799 – 2426-AV264732, adult male testis (DH10B), 4930502J08 3' 3'-BB759916, melanocyte, G270123D12 3' 3'-AV264732, adult male testis (DH10B), 4930502J08 3' 3'-BB730118, 8 cells embryo, E860116K03 3' 3'-BB760084, melanocyte, G270124K12 3' 3'-BB725787, 8 cells embryo, E860007M16 3' 3'-BB713373, 2 cells egg, B020045N21 3' 3'-BB727818, 8 cells embryo, E860029G22 3' 3'-BB732395, 12 days embryo whole body, E970024I12 3' 3'-BB678601, 16 days embryo head, 4121404G22 3' 3'-BB738719, 6 days neonate spleen, F430109A13 3' 3'-BB725966, 8 cells embryo, E860009P06 3' 3'-BB798298, adult male intestinal mucosa, G630012N23 3' 3'-BB808558, adult male intestinal mucosa, G630089N18 3' 3'-BB726922, 8 cells embryo, E860020P10 3' 3'-BB801972, adult male intestinal mucosa, G630041B06 3'

ing FANTOM2 and the proximity of cataloging the complete transcriptome.

Summary and Future Implications

The analysis of KIF functions is the most fundamental issue in elucidating the mechanism of intracellular transport. Related to this goal, how KIFs recognize and bind to specific cargo is another important question remaining to be solved.

Recent studies have begun to reveal that KIFs use scaffolding and adaptor protein complexes for this purpose (Nakagawa et al. 2000; Setou et al. 2000, 2002; Verhey et al. 2001). Another question that should be solved is how the cell determines the direction of transport and regulates KIF function.

The answers for these basic cell biological questions should be promoted using FANTOM resources of the mouse transcriptome. The FANTOM2 library will set the standard by serving as an encyclopedia for the future analysis of all transcribed molecules.

METHODS

Identification of All KIFs Contained in FANTOM2

We have screened for KIFs by using Pfam, InterPro, and Gene Ontology domain searches and auto-annotation and annotation by assigned curators from The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase II Team, 2002. The screen was confirmed by comprehensive BLASTN and TBLASTN searches using nucleotide and protein sequences of all KIFs. Results obtained by each method were recorded and compared (Table 1). KIFs with transcripts in FANTOM2 are indicated by a yellow-green underline in Figure 5.

Phase I Clones and EST Analysis

KIFs not found in the FANTOM2 data set were searched for in the Phase I database and the GenBank EST (Expression Sequenced Tag) database. Full-length, 5' end, and 3' end representative transcript sequences were used for BLASTN searches in the Phase I data set along with BLASTN searches in the GenBank mouse and human EST

databases. All ESTs in the Phase I set with a homology higher than 92% were recorded and are shown in Table 2. KIFs found only in the Phase I set are indicated by a black underline in Figure 5.

Sequence State Comparisons

All clones were compared with previous KIF sequences deposited in GenBank. The state of each clone was checked by comparison with full-length sequences deposited in GenBank and by a manual inspection of the deduced amino acid sequence. Clones containing the starting methionine residue and an in-frame stop downstream of the defined motor domain motif were considered full-length clones.

The KIF motor domain was defined by the following criteria: conservation of upstream p-loop motifs and a switch 2 sequence approximately 150–200 amino acid residues downstream, a YXXXXDLL motif where X is any amino acid and a switch 1 motif located between p-loop and switch 2 (Kikkawa et al. 2000). In addition to the ATP-binding motifs described above, the microtubule-binding motifs were also considered.

Splice Variant Identification

Clone ID's, library of origination and clone state in reference to full-length sequence including splice variants were recorded. The KIF clones as shown in Table 1, derived from analysis of all hits resulting from the screen described above. Splice variants were identified by comparison of nucleic and amino acid sequences of GenBank deposits and FANTOM2 clones. Alternative splicing was confirmed by the observation of different exons existing in proximal genomic sequences obtained from NCBI mouse genomic sequences. Novel exons were identified by examining intron sequences starting with nucleic sequences GT and ending with AG. Validation of the isoform was conducted by searching for ESTs encoding the splice form. Regions specific to respective isoforms were in reference to the number of ESTs in the NCBI mouse EST database. The number of supporting ESTs for respective splice forms was noted.

Phylogenetic Analysis

Figure 5 was reproduced with permission from the Proceedings of the National Academy of Science, U.S.A. 98(13) 7004–7011, 2001 (Miki et al. 2001). Briefly, the phylogenetic analysis was conducted by using the amino acid motor domain sequence of representative transcripts from all 45 loci in human and mouse, along with all representative KIF transcripts from *D. melanogaster*, *C. elegans*, and *S. cerevisiae*. Maximum parsimony was calculated (Tanaka et al. 1995) and the phylogram was drawn by TreeViewPPC (Page 1996). Bootstrap values were assessed by 10,000 random samplings. Classification of all KIFs was done as described previously (Hirokawa 1998).

ACKNOWLEDGMENTS

The authors are deeply in debt to other members of The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase II Team, 2002 and the Hirokawa lab. This work was funded by the Center of Excellence Grant-in-Aid from the Ministry of Education, Science, Sports, Culture and Technology of Japan to N. Hirokawa.

REFERENCES

Aizawa, H., Sekine, Y., Takemura, R., Zhang, Z., Nangaku, M., and Hirokawa, N. 1992. Kinesin family in murine central nervous system. *J. Cell Biol.* **119**: 1287–1296.

- Brendza, R.P., Serbus, L.R., Duffy, J.B., and Saxton, W.M. 2000. A function for kinesin I in the posterior transport of oskar mRNA and Staufen protein. *Science* **289**: 2120–2122.
- Guillaud, L., Setou, M., and Hirokawa, N. 2003. KIF17 dynamics and regulation of NR2B trafficking in hippocampal neurons. *J. Neurosci.* **23**: 131–140.
- Ha, M.J., Yoon, J., Moon, E., Lee, Y.M., Kim, H.J., and Kim, W. 2000. Assignment of the kinesin family member 4 genes (KIF4A and KIF4B) to human chromosome bands Xq13.1 and 5q33.1 by in situ hybridization. *Cytogenet. Cell Genet.* **88**: 41–42.
- Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405**: 283–284.
- Hirokawa, N. 1996. Organelle transport along microtubules—the role of KIFs. *Trends Cell Biol.* **6**: 135–141.
- . 1998. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* **279**: 519–526.
- Hirokawa, N., Noda, Y., and Okada, Y. 1998. Kinesin and dynein superfamily proteins in organelle transport and cell division. *Curr. Opin. Cell Biol.* **10**: 60–73.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**: 685–690.
- Kikkawa, M., Okada, Y., and Hirokawa, N. 2000. 15 Å resolution model of the monomeric kinesin motor, KIF1A. *Cell* **100**: 241–252.
- Kikkawa, M., Sablin, E.P., Okada, Y., Yajima, H., Fletterick, R.J., and Hirokawa, N. 2001. Switch-based mechanism of kinesin motors. *Nature* **411**: 439–445.
- Kim, A.J. and Endow, S.A. 2000. A kinesin family tree. *J. Cell Sci.* **113**: 3681–3682.
- Lander, E.S., Linton, L.M., Birren, B., Nussbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Miki, H., Setou, M., Kaneshiro, K., and Hirokawa, N. 2001. All kinesin superfamily protein, KIF, genes in mouse and human. *Proc. Natl. Acad. Sci.* **98**: 7004–7011.
- Nakagawa, T., Tanaka, Y., Matsuoka, E., Kondo, S., Okada, Y., Noda, Y., Kanai, Y., and Hirokawa, N. 1997. Identification and classification of 16 new kinesin superfamily (KIF) proteins in mouse genome. *Proc. Natl. Acad. Sci.* **94**: 9654–9659.
- Nakagawa, T., Setou, M., Seog, D., Ogasawara, K., Dohmae, N., Takio, K., and Hirokawa, N. 2000. A novel motor, KIF13A, transports mannose-6-phosphate receptor to plasma membrane through direct interaction with AP-1 complex. *Cell* **103**: 569–581.
- Nakajima, K., Takei, Y., Tanaka, Y., Nakagawa, T., Nakata, T., Noda, Y., Setou, M., and Hirokawa, N. 2002. Molecular motor KIF1C is not essential for mouse survival and motor-dependent retrograde Golgi Apparatus-to-endoplasmic reticulum transport. *Mol. Cell. Biol.* **22**: 866–873.
- Nangaku, M., Sato-Yoshitake, R., Okada, Y., Noda, Y., Takemura, R., Yamazaki, H., and Hirokawa, N. 1994. KIF1B, a novel microtubule plus end-directed monomeric motor protein for transport of mitochondria. *Cell* **79**: 1209–1220.
- Okada, Y., Yamazaki, H., Sekine-Aizawa, Y., and Hirokawa, N. 1995. The neuron-specific kinesin super family protein KIF1A is a unique monomeric motor for anterograde axonal transport of synaptic vesicle precursors. *Cell* **87**: 769–780.
- Okamoto, S., Matsushima, M., and Nakamura, Y. 1998. Identification, genomic organization, and alternative splicing of KNSL 3, a novel human gene encoding a kinesin-like protein. *Cytogenet. Cell Genet.* **83**: 25–29.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Olivier, M., Aggarwal, A., Allen, J., Almendras, A.A., Bajorek, E.S., Beasley, E.M., Brady, S.D., Bushard, J.M., Bustos, V.L., Chu, A., et al. 2001. A high-resolution radiation hybrid map of the human genome draft sequence. *Science* **291**: 1298–1302.
- Page, R.D. 1996. TreeView: An application to display phylogenetic trees on personal computers. *Comput. Applic. Biosci.* **12**: 357–358.
- Piddini, E., Schmid, J.A., de Martin, R., and Dotti, C.G. 2001. The Ras-like GTPase Gem is involved in cell shape remodelling and interacts with the novel kinesin-like protein KIF9. *EMBO J.* **20**: 4076–4087.
- Reddy, A.S.N. and Day, I.S. 2001. Kinesins in the *Arabidopsis*

- genome: A comparative analysis among eukaryotes. *BMC Genomics* **2**: 2–14.
- Setou, M., Nakagawa, T., Seog, D.H., and Hirokawa, N. 2000. Kinesin superfamily motor protein KIF17 and mLin-10 in NMDA receptor-containing vesicle transport. *Science* **288**: 1796–1802.
- Setou, M., Seog, D.H., Tanaka, Y., Kanai, Y., Takei, Y., Kawagishi, M., and Hirokawa, N. 2002. Glutamate-receptor-interacting protein GRIP1 directly steers kinesin to dendrites. *Nature* **417**: 83–87.
- Sharp, D.J., Rogers, G.C., and Scholey, J.M. 2000. Microtubule motors in mitosis. *Nature* **407**: 41–47.
- Tanaka, Y., Zhang, Z., and Hirokawa, N. 1995. Identification and molecular evolution of new dynein-like protein sequences in rat brain. *J. Cell Sci.* **108**: 1883–1893.
- Vale, R.D. and Fletterick, R.J. 1997. The design plan of kinesin motors. *Annu. Rev. Cell Develop. Biol.* **13**: 745–777.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Verhey, K.J., Meyer, D., Deehan, R., Blenis, J., Schnapp, B.J., Rapoport, T.A., and Margolis, B. 2001. Cargo of kinesin identified as JIP scaffolding proteins and associated signaling molecules. *J. Cell Biol.* **152**: 959–970.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Yamazaki, H., Nakata, T., Okada, Y., and Hirokawa, N. 1995. KIF3A/B: a heterodimeric kinesin superfamily protein that works as a microtubule plus end-directed motor for membrane organelle transport. *J. Cell Biol.* **130**: 1387–1399.
- Yang, Z., Hanlon, D.W., Marszalek, J.R. and Goldstein, L.S. 1997. Identification, partial characterization, and genetic mapping of kinesin-like protein genes in mouse. *Genomics* **45**: 123–131.
- Yang, Z., Roberts, E.A., and Goldstein, L.S. 2001. Functional analysis of mouse C-terminal kinesin motor KifC2. *Mol. Cell. Biol.* **21**: 2463–2466.
- Zhao, C., Takita, J., Tanaka, Y., Setou, M., Nakagawa, T., Takeda, S., Yang, H.W., Terada, S., Nakata, T., Takei, Y., et al. 2001. Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1B β . *Cell* **105**: 587–597.

Received November 25, 2002; accepted in revised form March 24, 2003.